

F. Altarelli,¹ A. Braunstein,² J. Realpe-Gomez,² and R. Zecchina²

¹ISI Foundation, Viale Settimio Severo 65, 10133 Torino, Italy

²Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino, Italy

(Dated: April 27, 2009)

Finding the optimal assignment in budget-constrained auctions is a combinatorial optimization problem with many important applications, a notable example being the sale of advertisement space by search engines (in this context the problem is often referred to as the off-line *AdWords* problem). Based on the cavity method of statistical mechanics, we introduce a message passing algorithm that is capable of solving efficiently random instances of the problem extracted from a natural distribution, and we derive from its properties the phase diagram of the problem. As the control parameter (average value of the budgets) is varied, we find two phase transitions delimiting a region in which long-range correlations arise.

I. INTRODUCTION

A. Problem definition

We consider the following model of budget-constrained auctions, inspired from the sale of advertisement space on internet search engines: a set of *advertisers* $a \in \{1, \dots, N_a\}$ are interested in appearing on the results pages of the searches relative to some *keywords* $k \in \{1, \dots, N_k\}$. Advertiser a offers to pay a *bid* $w_{ka} \in \mathbb{R}^+$ in order to appear on the results page each time that keyword k is searched. We assume that all the bids are expressed *before* the auction begins. In this setting, advertiser a doesn't know how many auctions he will end up winning, and therefore how much he will spend. In order to encourage the advertisers to make more bids without risking to spend too much, each advertiser a can also specify a *budget* $B_a \in \mathbb{R}^+$ which is the maximum sum that he is willing to pay in a given period of time.

The general problem we want to solve is the following: given the sets of bids $\{w_{ka}\}$ and of budgets $\{B_a\}$, what is the assignment of each keyword k to some advertiser a which maximizes the total revenue for the search engine? We can represent a possible assignment by introducing the binary variable x_{ka} which will take the value 1 if keyword k is assigned to advertiser a and 0 otherwise. The constraint that each keyword be assigned to one and only one advertiser then takes the form

$$\sum_a x_{ka} = 1 \quad (\forall k) \quad (1)$$

while the budget constraint takes the form

$$\sum_k x_{ka} w_{ka} \leq B_a \quad (\forall a) \quad (2)$$

and the quantity we want to maximize is the revenue of the search engine

$$\sum_a \sum_k x_{ka} w_{ka} \quad (3)$$

where $x \equiv \{x_{ka}\}$. In order to maximize the revenue, we shall make it possible to sell keywords at a discounted price when this allows to saturate the budget of some advertiser which would otherwise remain unsaturated, and define the revenue as

$$R(x) = \sum_a \min \left[B_a, \sum_k x_{ka} w_{ka} \right]. \quad (4)$$

Clearly, this problem is a variant of the weighted bipartite matching in which the budget constraint has been added. It is also a special case of the general linear Resource Allocation Problem on binary variables, which consists in finding $\{x_i\} \in \{0, 1\}^n$ that maximizes $\sum_i a_i x_i$ under the constraints $\sum_j b_{ij} x_j \leq c_i$ ($i = 1, \dots, m$).

B. Summary of known results

Both the on-line and the off-line versions of the AdWords problem have been the object of considerable attention in recent years. The off-line version is NP-hard even if there are only 2 advertisers [1]. An exact algorithm to solve it with time complexity $O(N_a 4^{N_k})$ and an approximate algorithm based on Integer Programming relaxation with an approximation ratio¹ $e/(e-1)$ (on the revenue) are presented in [2]. An improvement over this result is provided by [3], where an algorithm with approximation ratio $3/2$ is introduced. An approximate algorithm for the on-line version of the problem with competitive ratio² $e/(e-1)$ is introduced in [4].

These results are concerned with the *worst-case* analysis of the performance (scaling of time and approximation ratio) of the algorithms. In what follows, we shall be interested in a different question: is there an algorithm which has a good performance *on average* for some given ensemble of instances of the problem? The ensemble of instances will be specified by defining the distributions of the bids and of the budgets.

II. STATISTICAL MECHANICS ANALYSIS OF THE PROBLEM

A. Factor graph representation and probability marginals

We want to consider the problem of budget-constrained auctions as a statistical mechanics system, as done for example in [5]. The configurations $x \equiv \{x_{ka}\}$ will represent the assignments ($x_{ka} = 1$ if keyword k is assigned to advertiser a and 0 otherwise), and the energy function $E(x)$ will represent the portion of the advertisers budgets which remains unspent:

$$E(x) = \sum_a E_a(x_a) \equiv \sum_a \max \left[0, B_a - \sum_{k \in \partial a} w_{ka} x_{ka} \right] \quad (5)$$

where $\partial a \equiv \{k | w_{ka} > 0\}$, and $x_a \equiv \{x_{ka} | k \in \partial a\}$.

We want to include the constraint (1) that each keyword be assigned to one and only one advertiser as a *hard constraint*. We therefore write the Boltzmann-Gibbs distribution at inverse temperature β as:

$$P_\beta(x) = \frac{1}{Z(\beta)} \prod_a e^{-\beta E_a(x_a)} \prod_k \mathbb{1} \left[\sum_a x_{ka} = 1 \right] \quad (6)$$

where the partition function $Z(\beta)$ is a normalization.

This factorization of the probability distribution into local terms involving only some variables and corresponding to the assignment constraints (1) and the energy terms (5) suggests a *factor graph* representation [6] in which an instance of the problem is associated to a bipartite graph $\mathcal{G} = (F, V; E)$ where $F = \{a\} \cup \{k\}$ is the set of *function nodes* corresponding to the constraints, $V = \{x_{ka}\}$ is the set of *variable nodes* and E is the set of edges (a, x_{ka}) and (k, x_{ka}) such that $w_{ka} > 0$. In the following we shall assume that as $N_a, N_k \rightarrow \infty$, the average number of bids expressed by each advertiser remains *finite*. This kind of diluted systems can be successfully treated with the *cavity method* of statistical mechanics [7, 8], originally applied to optimization in [9, 10].

Given an instance of the problem $\mathcal{G} = (F, V; E)$, let us consider a *modified* instance $\mathcal{G}^{(k)}$ in which the function node k is removed, and another instance $\mathcal{G}^{(a)}$ in which the function node a is removed (the rest of these graphs being identical to \mathcal{G}). The name *cavity* refers to the missing node in $\mathcal{G}^{(k)}$ and in $\mathcal{G}^{(a)}$. We shall now make an assumption which goes under the name of *Replica Symmetric approximation* (RS), and which will be validated *ex-post*: we shall assume that variables x_{ka} and x_{jb} that are *far away* on the graph are uncorrelated, i.e. that their marginal distribution factorizes: $P(x_{ka}, x_{jb}) = P(x_{ka})P(x_{jb})$. Under this assumption, when a function node is removed from a diluted random factor graph, the variables connected to the removed node become far away from each other and therefore uncorrelated. Let us then consider the marginal distribution $P^{(a)}(x_{ka})$ for variable x_{ka} in the absence of node a : the value of x_{ka} is

¹ The *approximation ratio* is defined as the ratio between the best possible revenue and the revenue achieved by the algorithm in the worst case instance.

² The *competitive ratio* is defined as the ratio between the best possible revenue and the revenue achieved by the algorithm for the worst case sequence of choices of the on-line instance.

influenced only by the constraint in node k , so that

$$P_{ka}^{(a)}(x_{ka}) = \mathcal{N}_{ka}^{(a)} \sum_{\{x_{kb} | b \in \partial k \setminus a\}} \mathbb{1} \left[x_{ka} + \sum_{b \in \partial k \setminus a} x_{kb} = 1 \right] P(\{x_{kb} | b \in \partial k \setminus a\}) \quad (7)$$

where $\mathcal{N}_{ka}^{(a)}$ is a normalization factor, $\partial k \equiv \{a | w_{ka} > 0\}$ and the sum is over the $2^{|\partial k|-1}$ possible configurations of the variables $\{x_{kb} | b \in \partial k \setminus a\}$. In the limit $N, M \rightarrow \infty$, the joint marginal probability $P(\{x_{kb} | b \in \partial k \setminus a\})$ appearing on the right hand side can be replaced with the *cavity* joint marginal probability $P^{(k)}(\{x_{kb} | b \in \partial k \setminus a\})$ (this corresponds to a self-consistency assumption on the cavity marginal distributions for systems with (N_a, N_k) function nodes and systems with $(N_a - 1, N_k)$ or $(N_a, N_k - 1)$ function nodes, as explained in detail in [7, 8]). But in the absence of node k , the variables x_{kb} are far away in the graph, and the RS assumption implies that the joint marginal distribution factorises:

$$P_{ka}^{(a)}(x_{ka}) = \mathcal{N}_{ka}^{(a)} \sum_{\{x_{kb} | b \in \partial k \setminus a\}} \mathbb{1} \left[x_{ka} + \sum_{b \in \partial k \setminus a} x_{kb} = 1 \right] \prod_{b \in \partial k \setminus a} P_{kb}^{(k)}(x_{kb}) . \quad (8)$$

With the same argument we obtain

$$P_{ka}^{(k)}(x_{ka}) = \mathcal{N}_{ka}^{(k)} \sum_{\{x_{ja} | j \in \partial a \setminus k\}} e^{-\beta E_a(x_a)} \prod_{j \in \partial a \setminus k} P_{ja}^{(a)}(x_{ja}) . \quad (9)$$

Since the variables are binary, the cavity marginal distributions have only one parameter and we shall represent them in terms of cavity *fields* H_{ka} and U_{ka} :

$$\frac{P_{ka}^{(k)}(0)}{P_{ka}^{(k)}(1)} = e^{-\beta H_{ka}} , \quad \frac{P_{ka}^{(a)}(0)}{P_{ka}^{(a)}(1)} = e^{-\beta U_{ka}} . \quad (10)$$

In terms of the cavity fields, the self-consistent equations for the cavity marginals (8, 9) become:

$$\begin{aligned} e^{-\beta U_{ka}} &= \frac{\sum_{b \in \partial k \setminus a} P_{kb}^{(k)}(1) \prod_{c \in \partial k \setminus \{a, b\}} P_{kc}^{(k)}(0)}{\prod_{b \in \partial k \setminus a} P_{kb}^{(k)}(0)} \\ &= \sum_{b \in \partial k \setminus a} e^{\beta H_{kb}} , \end{aligned} \quad (11)$$

$$\begin{aligned} e^{-\beta H_{ka}} &= \frac{\sum_{x_{a \setminus k}} e^{-\beta E_a(0, x_{a \setminus k})} \prod_{j \in \partial a \setminus k} P_{ja}^{(a)}(x_{ja})}{\sum_{x_{a \setminus k}} e^{-\beta E_a(1, x_{a \setminus k})} \prod_{j \in \partial a \setminus k} P_{ja}^{(a)}(x_{ja})} \\ &= \frac{\sum_{x_{a \setminus k}} e^{-\beta [E_a(0, x_{a \setminus k}) - x_{a \setminus k} \cdot U_{a \setminus k}]}}{\sum_{x_{a \setminus k}} e^{-\beta [E_a(1, x_{a \setminus k}) - x_{a \setminus k} \cdot U_{a \setminus k}]}} \end{aligned} \quad (12)$$

where $x_{a \setminus k} \equiv \{x_{ja} | j \in \partial a \setminus k\}$, $U_{a \setminus k} \equiv \{U_{ja} | j \in \partial a \setminus k\}$ and $x_{a \setminus k} \cdot U_{a \setminus k} \equiv \sum_{j \in \partial a \setminus k} x_{ja} U_{ja}$, and where

$$E_a(x_{ka}, x_{a \setminus k}) = \max \left[0, B_a - x_{ka} w_{ka} - \sum_{j \in \partial a \setminus k} x_{ja} w_{ja} \right] . \quad (13)$$

We can solve the system of coupled equations (11) and (12) by iteration: we start with some random values for the cavity fields $\{H_{ka}^0\}$ and compute $\{U_{ka}^1\}$ using (11) with the values of $\{H_{ka}^0\}$; then we compute $\{H_{ka}^1\}$ using (12) and the values of $\{U_{ka}^1\}$; we iterate this procedure until the values $\{U_{ka}^t\}$ and $\{H_{ka}^t\}$ are a fixed point of the system (11) and (12) within some prescribed numerical accuracy.

B. Zero temperature limit of the probability marginals

As we mentioned before, the optimal assignments of a given instance of the problem will correspond to the configurations with lowest energy. In order to select them, we shall take the zero temperature limit $\beta \rightarrow \infty$. When

doing this, we must keep in mind that in case there are multiple optimal assignments, some of the variables will take a unique value in all of them, and in physical terms they will be completely “polarized”, while other variables will be able to take different values in different optimal assignments. In order to take this into account, we are going to assume that as $\beta \rightarrow \infty$ the cavity fields scale as

$$H_{ka} = h_{ka} + \beta^{-1} \xi_{ka} + o(\beta^{-1}), \quad U_{ka} = u_{ka} + \beta^{-1} \eta_{ka} + o(\beta^{-1}) \quad (14)$$

where h_{ka} and u_{ka} are known as *hard* fields and ξ_{ka} and η_{ka} are known as *soft* fields. The soft fields become crucial for those variables for which $h_{ka} + u_{ka} = 0$ (these variables are precisely those that are not fully polarized), and they are irrelevant for the other variables.

Assuming the scaling (14) and substituting it into the equations for the cavity fields (11) and (12) we obtain, by comparing the terms of $O(1)$ and those of $O(\beta^{-1})$:

$$u_{ka} = - \max_{b \in \partial k \setminus a} h_{kb}, \quad (15)$$

$$\eta_{ka} = - \log \left[\sum_{\{b \in \partial k \setminus a \mid h_{kb} = -u_{ka}\}} e^{\xi_{kb}} \right] \quad (16)$$

and

$$h_{ka} = g^*(B_a) - g^*(B_a - w_{ka}), \quad (17)$$

$$\xi_{ka} = \log \left[\sum_{x_{a \setminus k} \in \{x_{a \setminus k}^*(B_a - w_{ka})\}} e^{x_{a \setminus k} \cdot \eta_{a \setminus k}} \right] - \log \left[\sum_{x_{a \setminus k} \in \{x_{a \setminus k}^*(B_a)\}} e^{x_{a \setminus k} \cdot \eta_{a \setminus k}} \right] \quad (18)$$

where

$$g^*(z) \equiv \min_{x_{a \setminus k}} g(z, x_{a \setminus k}), \quad (19)$$

$$g(z, x_{a \setminus k}) \equiv \max [0, z - x_{a \setminus k} \cdot w_{a \setminus k}] - x_{a \setminus k} \cdot u_{a \setminus k} \quad (20)$$

are functions of the hard cavity fields $\{u_{ja}\}$ even though we don't write explicitly this dependency, and where the two sums in (18) are restricted to local configurations $x_{a \setminus k}^*(B_a - w_{ka})$ such that $g(B_a - w_{ka}, x_{a \setminus k}^*(B_a - w_{ka})) = g^*(B_a - w_{ka})$ in the first sum and $x_{a \setminus k}^*(B_a)$ such that $g(B_a, x_{a \setminus k}^*(B_a)) = g^*(B_a)$ in the second sum.

It is possible to give a simple physical interpretation of the hard and soft fields by considering the following quantities:

$$\varepsilon_{ka}^{(a)}(x_{ka}) = \frac{\sum_{\{x_{jb} \mid (jb) \neq (ka)\}} E^{(a)}(x) e^{-\beta E^{(a)}(x)}}{\sum_{\{x_{jb} \mid (jb) \neq (ka)\}} e^{-\beta E^{(a)}(x)}}, \quad (21)$$

$$n_{ka}^{(a)}(x_{ka}) = e^{\beta \varepsilon_{ka}^{(a)}(x_{ka})} \sum_{\{x_{jb} \mid (jb) \neq (ka)\}} e^{-\beta E^{(a)}(x)} \quad (22)$$

where $E^{(a)}(x)$ is the energy of configuration x on the cavity graph $\mathcal{G}^{(a)}$. Clearly, $\varepsilon_{ka}^{(a)}(x_{ka})$ is the average value of the energy (at temperature β) on the cavity graph $\mathcal{G}^{(a)}$ when the variable x_{ka} is constrained to take a given value, while $\log(n_{ka}^{(a)}(x_{ka}))$ is the value of the entropy under the same conditions. When $\beta \rightarrow \infty$, $\varepsilon_{ka}^{(a)}(x_{ka})$ tends to the minimum value of the cavity energy over the configurations with given x_{ka} , while $n_{ka}^{(a)}(x_{ka})$ tends to the number of configurations that realize this minimum.

The cavity bias for variable x_{ka} can be written as

$$e^{-\beta H_{ka}} = \frac{P_{ka}^{(a)}(0)}{P_{ka}^{(a)}(1)} = \frac{\sum_{\{x_{jb} \mid (jb) \neq (ka)\}} e^{-\beta E^{(a)}(x_{ka}=0, \{x_{jb}\})}}{\sum_{\{x_{jb} \mid (jb) \neq (ka)\}} e^{-\beta E^{(a)}(x_{ka}=1, \{x_{jb}\})}} = \frac{e^{-\beta \varepsilon_{ka}^{(a)}(0)} n_{ka}^{(a)}(0)}{e^{-\beta \varepsilon_{ka}^{(a)}(1)} n_{ka}^{(a)}(1)} \quad (23)$$

from which we identify

$$h_{ka} = \varepsilon_{ka}^{(a)}(0) - \varepsilon_{ka}^{(a)}(1), \quad (24)$$

$$\xi_{ka} = \log n_{ka}^{(a)}(1) - \log n_{ka}^{(a)}(0). \quad (25)$$

The hard field h_{ka} is then equal to the difference in the average cavity energies of configurations with $x_{ka} = 0$ relative to configurations with $x_{ka} = 1$, while the soft field ξ_{ka} is equal to (minus) the difference in the cavity entropies between the same configurations. The fields u_{ka} and η_{ka} associated to a keyword cavity (k) can be interpreted in the same manner.

A few remarks are in order. First, notice that the equations (15,17) for the hard fields $\{h_{ka}\}$ and $\{u_{ka}\}$ are *independent* of the soft fields $\{\xi_{ka}\}$ and $\{\eta_{ka}\}$. One can then split the computation into two parts: first, the fixed point of the recursion (15, 17) is found for the hard fields, giving $\{h_{ka}^*\}$ and $\{u_{ka}^*\}$; then, these values are substituted (as constants) into the recursion (16, 18), which is solved. This is usually faster and simpler than doing a parallel update.

The second remark concerns the absolute values and the signs of the cavity fields: it is easily seen that $g^*(z)$ is a continuous non-decreasing function of z , so that the fields h_{ka} are positive reals³ or zero; this in turn implies that the fields u_{ka} are negative reals or zero. In a way, this could be expected from the message-passing interpretation of the fields: the messages u_{ka} “express” the constraint that keyword k be assigned to only one advertiser, and therefore tend to bias the variables towards 0 (corresponding to a negative field), while the messages h_{ka} “express” the requirement that the energy be as low as possible compatibly with the budget constraint, and therefore tend to bias the variables towards 1 (corresponding to a positive field). Similarly, it can be seen that the messages ξ_{ka} must be positive or null, so that the messages η_{ka} must be negative or null (the argument however is less trivial and we shall omit it).

As a final remark, let us show that all these update rules can be computed efficiently. This is obvious for the equations (15) and (16), which involve finding the minimum among $c - 1$ numbers (where c is the connectivity of the keyword considered). By looking for the two largest values of h_{ka} for a given k , one can update all the u_{ka} and η_{ka} messages on the edges incident on k in linear time in c . For the equations (17) and (18) it is less obvious: these equations require finding the minimum over the local configurations $x_{a \setminus k}$ of a function $g(z, x_{a \setminus k})$, i.e. to compute it in 2^{c-1} points, where c can be very large as $N_a, N_k \rightarrow \infty$. However, this can also be computed efficiently. Let us consider the functions $g_n^*(z)$ obtained by minimizing $g(z, \{x_1, \dots, x_n\})$ over the first n variables, which we denote by x_i for simplicity. The functions $g_n^*(z)$ obey the recursion:

$$g_{n+1}^*(z) \equiv \min_{\{x_1, \dots, x_{n+1}\}} \left\{ \max \left[0, z - \sum_{i=1}^{n+1} x_i w_i \right] - \sum_{i=1}^{n+1} x_i u_i \right\} \quad (26)$$

$$= \min [g_n^*(z), g_n^*(z - w_{n+1}) - u_{n+1}] \quad (27)$$

with the initial condition

$$g_0^*(z) = \max[0, z]. \quad (28)$$

It is then possible to compute the function $g^*(z)$ by iterating the recursion (27). This can be done either approximately, discretizing the values of z , or exactly, by exploiting the fact that $g^*(z)$ is a function which has slope equal to either one or zero in any point, and that the number of points where it changes slope remains finite as the connectivity increases (this actually depends on the distribution of the values of the bids $\{w_{ka}\}$ and of the fields $\{u_{ka}\}$, but it is true for the distributions we shall consider). The sums appearing in the logarithms of (18) can be computed together with the function $g^*(z)$ in a similar way.

C. Average energy and entropy in terms of the fields

For a given instance, the average energy and the entropy at zero temperature can be computed as functions of the fields $\{h_{ka}\}, \{\xi_{ka}\}$ (or equivalently $\{u_{ka}\}, \{\eta_{ka}\}$) by writing in their definitions,

$$E = \sum_x P(x) E(x), \quad (29)$$

$$S = - \sum_x P(x) \log P(x) \quad (30)$$

³ Notice that this contrasts many other combinatorial optimization problems in which the energy, and therefore the hard fields, are discrete quantities.

the distribution $P(x)$ in terms of the values of the fields at the fixed point:

$$P(x) = \prod_a P_a(x_a) \prod_k P_k(x_k) \prod_{(ka)} P_{ka}(x_{ka})^{-1}, \quad (31)$$

where as before $x_a \equiv \{x_{ka} | k \in \partial a\}$, $x_k \equiv \{x_{ka} | a \in \partial k\}$, and

$$P_a(x_a) = \mathcal{N}_a^{-1} \exp \{-\beta [E_a(x_a) - x_a \cdot U_a]\}, \quad (32)$$

$$P_k(x_k) = \mathcal{N}_k^{-1} \exp \{\beta x_k \cdot H_k\} \mathbb{1} \left[\sum_{b \in \partial k} x_{kb} = 1 \right], \quad (33)$$

$$P_{ka}(x_{ka}) = \mathcal{N}_{ka}^{-1} \exp \{\beta x_{ka} (H_{ka} + U_{ka})\} \quad (34)$$

and where \mathcal{N}_a , \mathcal{N}_k and \mathcal{N}_{ka} are normalizations.

From these equations one obtains the average energy in the limit $\beta \rightarrow \infty$:

$$E = \sum_a \frac{\sum_{x_a \in \{x_a^*\}} \max [0, B_a - x_a \cdot w_a] e^{x_a \cdot \eta_a}}{\sum_{x_a \in \{x_a^*\}} e^{x_a \cdot \eta_a}} \quad (35)$$

with the same notations as in the previous paragraph: $u_a \equiv \{u_{ka} | k \in \partial a\}$ and similarly for η_a and w_a ; $x_a \cdot \eta_a \equiv \sum_{k \in \partial a} x_{ka} \eta_{ka}$ and similarly for $x_a \cdot u_a$ and $x_a \cdot w_a$; $g(B_a, x_a) = \max [0, B_a - x_a \cdot w_a] - x_a \cdot u_a$, which is defined similarly to equation (20), except that *all* the variables incident on a are considered instead of all but one; $g^*(B_a) = \min_{x_a} g(B_a, x_a)$; $\{x_a^*\}$ is the set of partial configurations x_a that realize $g^*(B_a)$. When all the $\{u_{ka}\}$ are non-zero, this expression simplifies greatly:

$$E = \sum_a \max [0, B_a - x_a^* \cdot w_a] \quad (36)$$

where x_a^* is *any* of the (possibly degenerate) partial configurations that minimize $g(B_a, x_a)$. Such a configuration, in turn, is easy to compute if all the $\{h_{ka}\}$ are different from zero, as can be seen by writing explicitly the minimization with respect to one of the variables, say x_{ka} :

$$\begin{aligned} \min_{x_a} g(B_a, x_a) &= \min \left\{ \min_{x_{a \setminus k}} \left[\max (0, B_a - x_{a \setminus k} \cdot w_{a \setminus k}) - x_{a \setminus k} \cdot u_{a \setminus k} \right], \right. \\ &\quad \left. \min_{x_{a \setminus k}} \left[\max (0, B_a - w_{ka} - x_{a \setminus k} \cdot w_{a \setminus k}) - x_{a \setminus k} \cdot u_{a \setminus k} - u_{ka} \right] \right\} \end{aligned} \quad (37)$$

$$= \min_{x_{a \setminus k}} \left[\max (0, B_a - x_{a \setminus k} \cdot w_{a \setminus k}) - x_{a \setminus k} \cdot u_{a \setminus k} \right] + \min(0, -h_{ka} - u_{ka}) \quad (38)$$

where the choice of the first term in the two-term minimum corresponds to $x_{ka}^* = 0$ while that of the second corresponds to $x_{ka}^* = 1$, provided $h_{ka} \neq 0$. Under the combined assumption $u_{ka} \neq 0 \neq h_{ka} \forall (ka)$ we have $x_{ka}^* = \theta(h_{ka} + u_{ka})$, where $\theta(x) = 1$ for $x > 0$ and $\theta(x) = 0$ otherwise.

The expression of the entropy is somewhat more complicated:

$$\begin{aligned} S &= \sum_a \left\{ \log \sum_{x_a \in \{x_a^*\}} e^{x_a \cdot \eta_a} - \frac{\sum_{x_a \in \{x_a^*\}} x_a \cdot \eta_a e^{x_a \cdot \eta_a}}{\sum_{x_a \in \{x_a^*\}} e^{x_a \cdot \eta_a}} \right\} + \\ &\quad + \sum_k \left\{ \log \sum_{a \in \partial k: h_{ka} = h_k^*} e^{\xi_{ka}} - \frac{\sum_{a \in \partial k: h_{ka} = h_k^*} \xi_{ka} e^{\xi_{ka}}}{\sum_{a \in \partial k: h_{ka} = h_k^*} e^{\xi_{ka}}} \right\} + \\ &\quad - \sum_{ka} \mathbb{1} [h_{ka} + u_{ka} = 0] \left\{ \log (1 + e^{\xi_{ka} + \eta_{ka}}) - \frac{(\xi_{ka} + \eta_{ka}) e^{\xi_{ka} + \eta_{ka}}}{1 + e^{\xi_{ka} + \eta_{ka}}} \right\} \end{aligned} \quad (39)$$

with the same notations as above and $h_k^* \equiv \max_{a \in \partial k} h_{ka}$.

Once the fixed-point fields $\{h_{ka}, \xi_{ka}\}$ (or equivalently $\{u_{ka}, \eta_{ka}\}$) are known, it is straightforward to compute the other set of fields via (15, 16) or via (17, 18), and then to compute the energy and entropy of the ground states via (35, 39).

A. Characterization of the random ensemble of instances

An instance of the problem can be represented as a bipartite graph $\mathcal{I} = (K, A; E)$ in which $K = \{1, \dots, N_k\}$ is the set of keywords, $A = \{1, \dots, N_a\}$ is the set of advertisers, and $E = \{1, \dots, N_e\}$ is the set of the edges (ka) such that $w_{ka} > 0$ (notice that this graph should not be confused with the factor graph introduced earlier). We assume the graph to be random with given N_a , N_k and N_e , and that in the thermodynamic limit $N_a, N_k, N_e \rightarrow \infty$ the average connectivities $c_a \equiv N_e/N_a$ and $c_k \equiv N_e/N_k$ both remain finite. We consider the (non-zero) bids w_{ka} to be iid random variables with uniform distribution in $[0, 1]$. Given the bids $\{w_{a1}, \dots, w_{ac}\}$ offered by an advertiser a with connectivity c , any budget smaller than $B_a^{\min} \equiv \min_j w_{aj}$ or larger than $B_a^{\max} \equiv \sum_j w_{aj}$ will be meaningless. We shall therefore consider a distribution of budgets *correlated* to the connectivities and to the bids, in such a way that each budget B_a is constrained to the interval $[B_a^{\min}, B_a^{\max}]$. The most “natural” distribution of budgets is obtained by conditioning the *uniform* distribution in $\bigotimes_a [B_a^{\min}, B_a^{\max}]$ to have a given sum, specified by a parameter $\bar{b} \in [0, 1]$:

$$\mathbb{P} [\{B_a\} | \bar{b}, \{w_{ka}\}] \propto \delta \left(\sum_a [B_a^{\min} + \bar{b} (B_a^{\max} - B_a^{\min})] - \sum_a B_a \right) \prod_a \theta(B_a - B_a^{\min}) \theta(B_a^{\max} - B_a). \quad (40)$$

As $N_a \rightarrow \infty$ the “reduced” budgets $b_a \equiv (B_a - B_a^{\min}) / (B_a^{\max} - B_a^{\min})$ become iid random variables with an exponential distribution in the interval $[0, 1]$ with parameter γ given by the (unique) solution of $1 + 1/(e^\gamma - 1) - 1/\gamma = \bar{b}$. This makes it very simple to generate large instances from this distribution. In the thermodynamic limit the average budget \bar{B} is related to the parameter \bar{b} by the linear relation:

$$\bar{B} = \frac{e^{-c_a}}{c_a} (e^{c_a} - 1 - c_a) (1 - \bar{b}) + \frac{c_a \bar{b}}{2}. \quad (41)$$

We shall denote this random ensemble as $\mathcal{E}(N_a, N_k, N_e, \bar{B})$ for finite-size instances and as $\mathcal{E}_\infty(c_a, c_k, \bar{b})$ in the thermodynamic limit. In the following sections, we shall discuss how the properties of \mathcal{E} and \mathcal{E}_∞ vary with these parameters. Of course, one can consider many other ensembles of instances. For example, one could consider an exponential distribution for the bids, or remove the correlation between bids and budgets. In fact, only the distribution of the *actual* bids and budgets can claim to be more interesting than the other possible ones, and the only motivation for our choice is that it is reasonable and simple.

B. Definition of the algorithm

We shall now define an algorithm, based on the analysis of the previous section, which finds efficiently optimal (or almost-optimal) configurations for random instances of the problem extracted from the distribution described in the previous paragraph, and which coincides with the zero-temperature Belief Propagation (BP) algorithm, with the cavity fields playing the role of messages, and the self-consistent equations for the cavity fields playing the role of update rules. We shall see that in order to find an optimal configuration of the problem it is sufficient to compute the hard fields, so the update equations for the messages will be (15) and (17), which we rewrite as $h_{ka} = \mathcal{H}_{ka}(u_{a \setminus k})$ and $u_{ka} = \mathcal{U}_{ka}(h_{k \setminus a})$. For the sake of concreteness the pseudo-code of the Belief Propagation algorithm is written in Algorithm III.1. The inputs are an off-line AdWords instance \mathcal{I} , an initial set of messages $\{h_{ka}^0\}$, a convergence criterion $\mathcal{C}(\{h_{ka}^t\}, \epsilon) \in \{\text{true, false}\}$, and a maximum number of iterations T .

This algorithm is successful in finding an optimal configuration only if the following two conditions are met: the messages must converge to a fixed point, and at the fixed point $h_{ka} + u_{ka}$ must be non-zero for all (ka) . As we shall see in the following section, these conditions are not always verified for random instances extracted from the random ensemble $\mathcal{E}(N_a, N_k, N_e, \bar{B})$ described in paragraph III A. However, we shall see that it is possible to modify the update equations (15) and (17) so that a solution is (almost) always found.

IV. NUMERICAL RESULTS

A. Convergence of the zero-temperature BP

We have implemented the Algorithm III.1 and used it to solve random instances from the distribution \mathcal{E} described in paragraph III A, with several values for the distribution parameters N_a , N_k , N_e and \bar{b} .

Algorithm III.1: ZERO-TEMPERATURE BP($\mathcal{I}, \{h_{ka}^0\}, \mathcal{C}, T$)

```

 $t \leftarrow 0$ 
while  $(-\mathcal{C}(\{h_{ka}^t\}, \epsilon))$  and  $(t < T)$ 
  do  $\left\{ \begin{array}{l} \text{for each } (ka) \in E \\ \text{do } u_{ka}^t \leftarrow \mathcal{U}_{ka}(h_{k \setminus a}^t) \end{array} \right.$ 
     $t \leftarrow t + 1$ 
    for each  $(ka) \in E$ 
      do  $h_{ka}^t \leftarrow \mathcal{H}_{ka}(u_{a \setminus k}^{t-1})$ 
  for each  $(ka) \in E$ 
    do  $u_{ka}^t \leftarrow \mathcal{U}_{ka}(h_{k \setminus a}^t)$ 
  if  $(\mathcal{C}(\{h_{ka}^t\}, \epsilon))$  and  $(h_{ka}^t + u_{ka}^t \neq 0 \ \forall (ka))$ 
    then  $\left\{ \begin{array}{l} \text{for each } ka \in E \\ \text{do } x_{ka} \leftarrow \theta(h_{ka}^t + u_{ka}^t) \end{array} \right.$ 
  else return (Undetermined)

```

In Figures 1 and 3 we study the convergence behavior of the algorithm as the budget distribution parameter \bar{b} is varied, while the other parameters are kept fixed: $N_a = 1000$, $N_k = 3000$, $N_e = 10000$. The algorithm is run with the following parameters: the initial values of the messages are set to $h_{ka}^0 = 0$ for all (ka) , the maximum number of iteration is $T = 2000$, the convergence criterion is

$$\mathcal{C}(\{h_{ka}^t\}, \epsilon) = \left\{ \sum_{(ka)} \left[|h_{ka}^t - h_{ka}^{t-1}|^2 + |u_{ka}^t - u_{ka}^{t-1}|^2 \right] < \epsilon^2 \right\} \wedge \left\{ \theta(h_{ka}^t + u_{ka}^t) = \theta(h_{ka}^{t-1} + u_{ka}^{t-1}), \forall (ka) \right\} \quad (42)$$

with $\epsilon = 10^{-5}$. A total of 31 736 instances are generated with values of \bar{b} in the range $[0.17, 0.45]$, and they are grouped according to the *actual* value of the average budget \bar{B} , in bins of amplitude 0.02 (recall that since the N_a individual budgets are extracted as iid random variables from a distribution with parameter \bar{b} , their average \bar{B} is also a random variable).

For a given bin, if the total number of samples is n and the number of samples for which convergence is obtained is m , the distribution for the probability (over the choice of the instance) of convergence p is computed as follows⁴:

$$\mathbb{P}[m|n, p] = \binom{n}{m} p^m (1-p)^{n-m} \Rightarrow \mathbb{P}[p|n, m] = (n+1) \binom{n}{m} p^m (1-p)^{n-m} \quad (43)$$

so that the expected value of the convergence probability \bar{p} and its standard deviation δp are:

$$\bar{p} = \frac{m+1}{n+2}, \quad (44)$$

$$\delta p = \sqrt{\frac{m+1}{n+2} \left(\frac{m+2}{n+3} - \frac{m+1}{n+2} \right)}. \quad (45)$$

In Figure 1 we show the results for \bar{p} and δp as a function of \bar{B} . For small values, $\bar{B} \lesssim 1.4$, the algorithm almost always converges, while for $1.4 \lesssim \bar{B} \lesssim 1.5$ there is a sharp decrease in the probability of convergence. For larger values, $1.5 \lesssim \bar{B}$, the probability of convergence gradually increases back to 1, reaching the value $\bar{p} \simeq 0.5$ for $\bar{B} \simeq 1.8$.

The equivalence of the convergence of our algorithm with the RS assumption together with the above results suggest a phase diagram (in terms of \bar{B}) with a RS phase for $\bar{B} < \bar{B}_0$, a Replica Symmetry Broken (RSB) phase for $\bar{B}_0 < \bar{B} < \bar{B}_1$ and again a RS phase for $\bar{B}_1 < \bar{B}$, where \bar{B}_0 and \bar{B}_1 are functions of the average connectivities c_a and c_k and are roughly equal, for $c_a = 10$ and $c_k = 10/3$, to 1.4 and 1.8 respectively. The existence of a phase transition at $\bar{B} \simeq 1.4$ is supported by the scaling of \bar{p} with the size of the instance (i.e. N_a) shown in Figure 2, which indicates that slightly above $\bar{B} = 1.4$ in the infinite size limit \bar{p} tends to zero. The transition at $\bar{B} = 1.8$ on the other hand

⁴ Assuming a uniform prior over p .

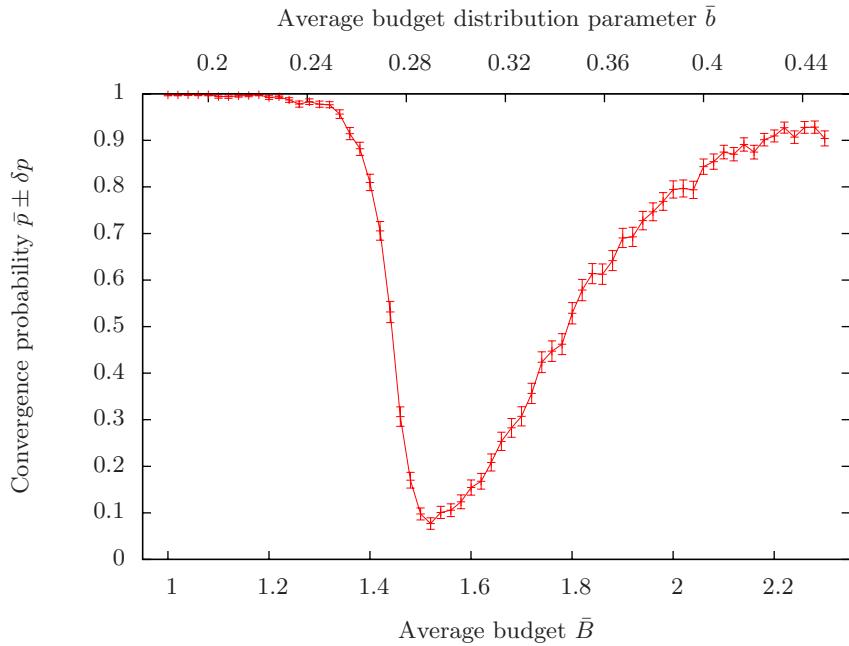


FIG. 1: Convergence of the algorithm for different values of the average budget \bar{B} . The other parameters of the instance distribution are as follows: $N_a = 1000$, $N_k = 3000$ and $N_e = 10000$ (the parameters for the algorithm are specified in the text). The samples are grouped according to their average budget \bar{B} in bins of amplitude 0.02 (the horizontal errorbars are not shown for simplicity); the corresponding values of the distribution parameter \bar{b} , computed according to (41), are shown on the upper scale. The number of samples in each bin is approximately 500.

does *not* appear to be sharp. This is not in contradiction with the suggested phase diagram, since the convergence probability on finite size instances might be affected by the presence of local structures where the fields do not reach convergence, but which do not contribute to the properties of the ensemble in the infinite size limit, as happens for example in the random $(2+p)$ -XORSAT. Indeed, we shall see that the analysis of the stability of the RS solution (in paragraph IV B) confirms this phase diagram.

It should be stressed that these results on convergence could, in principle, depend heavily on the choices of the convergence criterion $\mathcal{C}(\{h_{ka}^t\}, \epsilon)$, of the maximum number of iterations allowed T , and of the initial values of the messages h_{ka}^0 . We find that the choice of the convergence criterion has very limited impact, as long as it is a reasonable one. We experimented with several values of ϵ between 10^{-10} and 10^{-1} , and we also tried other possible criteria, such as the following two:

$$\mathcal{C}'(\{h_{ka}^t\}, \epsilon) = \left\{ \max_{(ka)} [\max(|h_{ka}^t - h_{ka}^{t-1}|, |u_{ka}^t - u_{ka}^{t-1}|)] < \epsilon \right\} \wedge \left\{ \theta(h_{ka}^t + u_{ka}^t) = \theta(h_{ka}^{t-1} + u_{ka}^{t-1}), \forall (ka) \right\}, \quad (46)$$

$$\mathcal{C}''(\{h_{ka}^t\}, \epsilon) = \left\{ \theta(h_{ka}^t + u_{ka}^t) = \theta(h_{ka}^{t-1} + u_{ka}^{t-1}) = \dots = \theta(h_{ka}^{t-4} + u_{ka}^{t-4}), \forall (ka) \right\}. \quad (47)$$

In all these cases, we find very similar results. We also tried an alternative for the initial condition: setting h_{ka}^0 to random values extracted uniformly in the interval $[0, 1]$. Again, this has very limited impact on the convergence behavior of the algorithm. Finally, we analyzed the possible effect of the cut-off on the number of iterations allowed. In Figure 3 we show the average number of iterations n_i required for convergence, together with the fraction of runs that converge and require less than 1000 iterations (i.e. half the maximum number of iterations allowed, T). By comparing the two plots, and assuming that the distribution of n_i is unimodal when $T \rightarrow \infty$, it seems clear that the effect of this cut-off is negligible for the samples considered.

In Figure 4, we see that for $\bar{B} \lesssim 1.5$ a large fraction of the variables are unfrozen, that is to say that the fields associated to them verify $h_{ka} + u_{ka} = 0$, so that the simple prescription used by Algorithm III.1 is unable to assign them. The existence of regions of values of \bar{B} where the messages often do not converge to a fixed point, or where the fixed point corresponds to a large number of unfrozen variables, constitutes a serious problem for the practical use of our algorithm. However, as we mentioned before, it is possible to “fix” it with two simple modifications of the update equations (15) and (17).

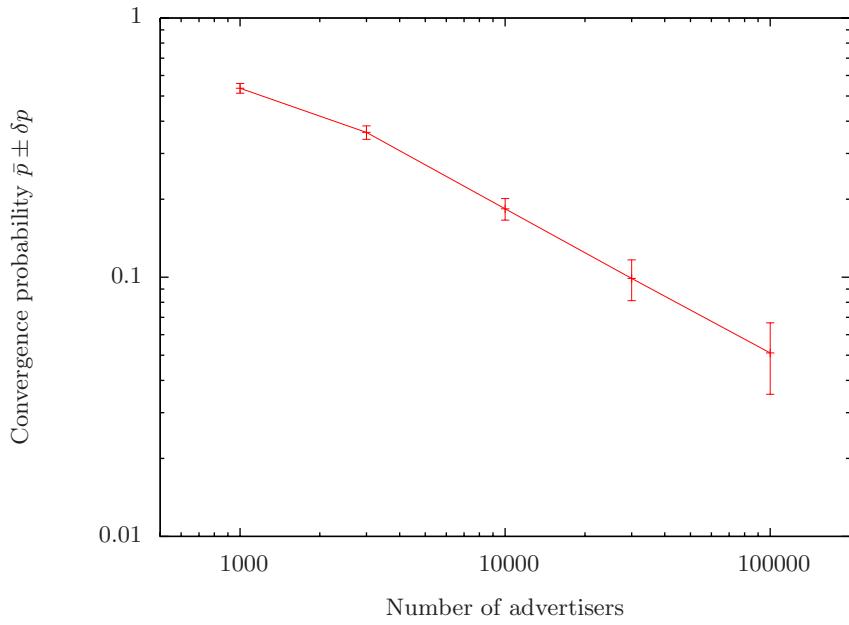


FIG. 2: Scaling of the convergence probability \bar{p} with the number of advertisers N_a for fixed $N_k/N_a = 3$, $N_e/N_a = 10$ and $\bar{B} = 1.44$. The data points are very well fitted by $\bar{p} \propto N_a^\nu$ with the exponent $\nu = -1/2$, indicating a sharp transition in the infinite size limit.

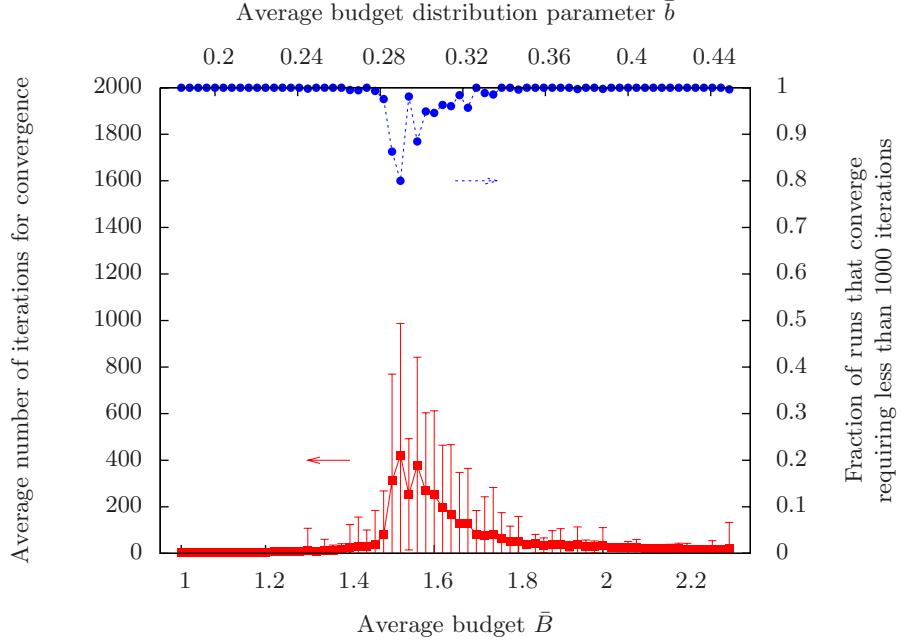


FIG. 3: Number of iterations required for convergence (same samples as in Figure 1). The bottom curve (left scale) shows the average number of iterations required for convergence and its standard error. The upper curve (right scale) shows the fraction of the samples that converge requiring less than 1000 iterations.

The probability of convergence \bar{p} of the algorithm is greatly improved if the update rule for the fields h_{ka}^t is modified to include a *reinforcement* term [11]. The modified rule takes the form:

$$h_{ka}^t \leftarrow \mathcal{H}_{ka}(u_{a \setminus k}^{t-1}) + \rho(h_{ka}^{t-1} + u_{ka}^{t-1}) \quad (48)$$

where ρ is a constant real number which we shall refer to as *reinforcement parameter*. In practice, the value taken by the new field is a linear combination of the value prescribed by the update equation (17) and the value of the *full* field

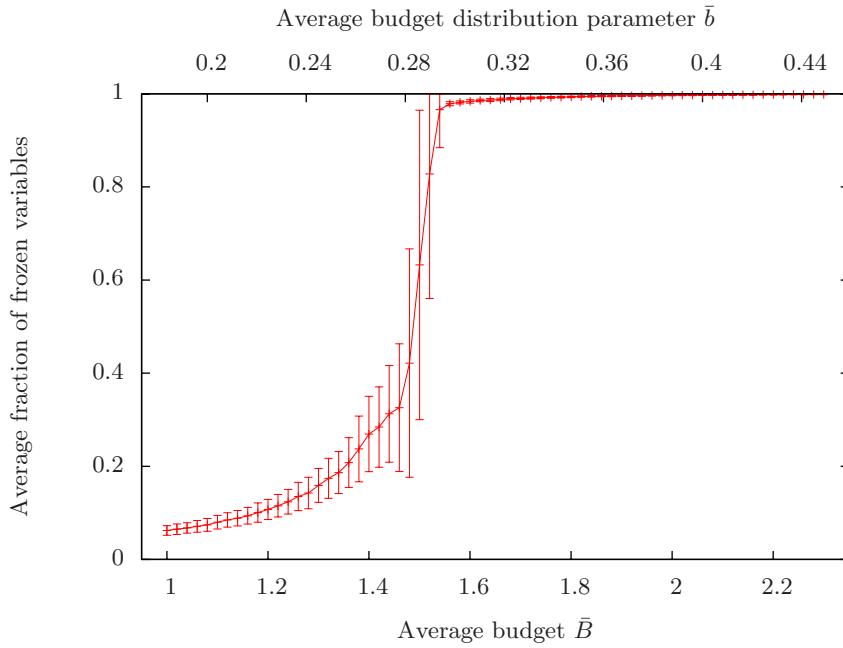


FIG. 4: Fraction of *frozen* variables, i.e. x_{ka} such that $h_{ka} + u_{ka} \neq 0$ (same samples as Figures 1 and 3).

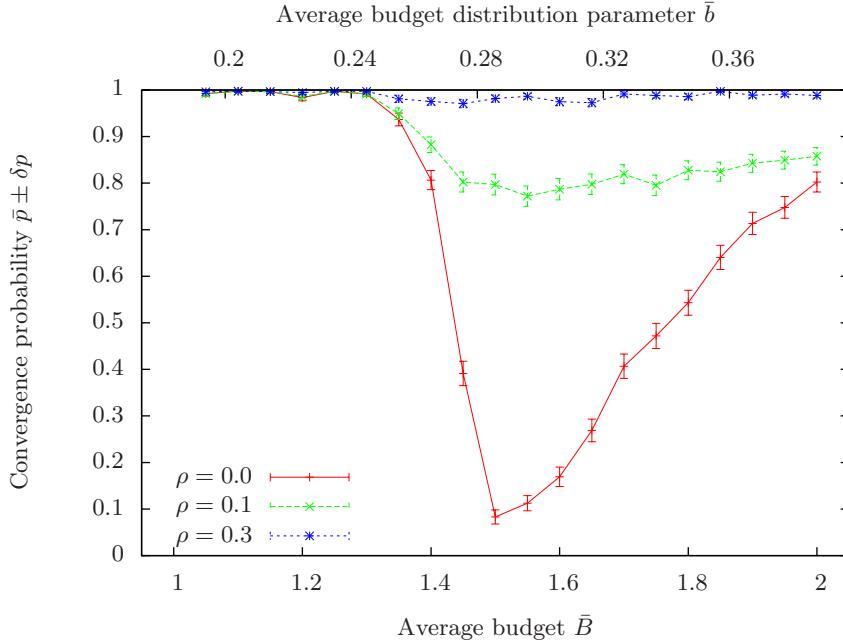


FIG. 5: Effect of reinforcement on the convergence of the algorithm. The parameters of the instances distribution are the same as in Figure 1, but a reinforcement term with parameter ρ is included in the update equation (48).

(as opposed to the cavity one) acting on the variable being updated, which is equal to $h_{ka}^{t-1} + u_{ka}^{t-1}$. The coefficient ρ is used to “tune” the strength of this term.

The results of this modification in the update rule are quite spectacular, as can be seen on Figure 5, where the probability of convergence \bar{p} is compared, for the *same* samples, with different values of the reinforcement parameter ρ . The parameters of the distribution of instances are the same as in Figures 1 – 4, i.e. $N_a = 1000$, $N_k = 3000$, $N_e = 10000$, and the total number of instances is 6 771. The parameters for the algorithm are as follows: the convergence criterion used is \mathcal{C}'' , defined in (47), the maximum number of iterations allowed is $T = 800$, and the initial values of the messages are $h_{ka}^0 = 0$ for all (ka) . For $\rho = 0.3$, approximately 99% of the samples converge within

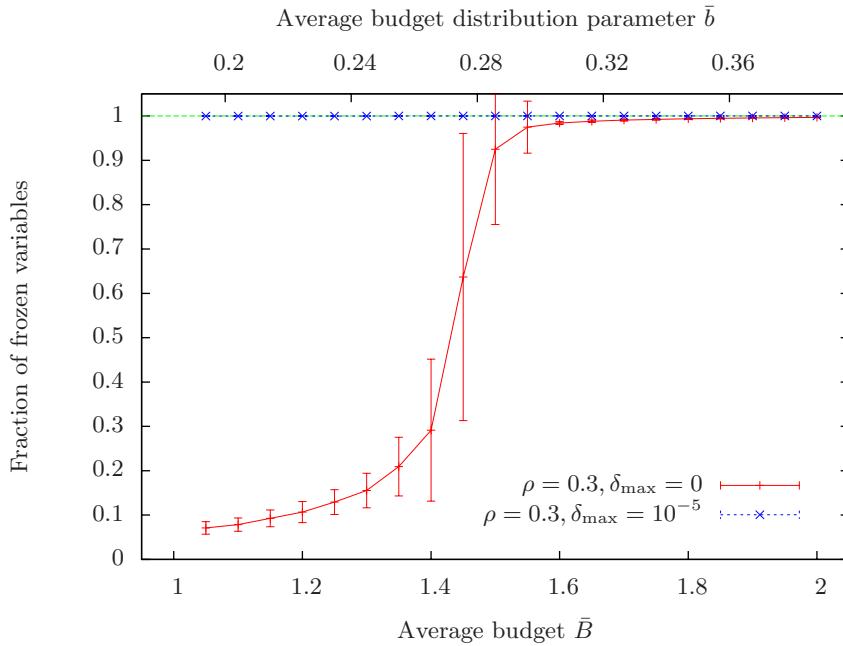


FIG. 6: Effect of random pinning fields on the numbers of frozen variables. The instances are the same as in Figure 5.

500 iteration.

The second problem of the original algorithm, namely that a large fraction of the variables are unfrozen for small values of \bar{B} , can also be easily solved, by adding a random *pinning field* to the update equations, which become (including the reinforcement term):

$$h_{ka}^t \leftarrow \mathcal{H}_{ka}(u_{a \setminus k}^{t-1}) + \rho(h_{ka}^{t-1} + u_{ka}^{t-1} - \delta_{ka}) + \delta_{ka}, \quad (49)$$

$$u_{ka}^t \leftarrow \mathcal{U}_{ka}(h_{k \setminus a}^t) + \delta_{ka} \quad (50)$$

where the pinning fields δ_{ka} are constant random real numbers, extracted from a uniform distribution in the interval $]0, \delta_{\max}]$. Notice that the pinning field also enters in the reinforcement term: since it affects both the updates of h_{ka} and u_{ka} , its effect is duplicated in the sum $h_{ka} + u_{ka}$, which must therefore be corrected as $h_{ka} + u_{ka} - \delta_{ka}$ (this also applies to the step functions defining the values of x_{ka} and the convergence criteria). Figure 6 shows the effect of the pinning field on the average number of frozen variables: all the variables become frozen, except at most 6 out of 10 000 in the samples considered.

With these two modifications (i.e. the introduction of a reinforcement term and of random pinning fields), our algorithm is almost always capable to find a low-energy configuration of the problem. However, the introduction of these “spurious” terms in the update equations could lead to an increase in the energy of the configuration obtained, relative to the true optimal energy of the instance. In order to verify that this effect is not significant, we have compared the energy of the configuration found by the algorithm using both reinforcement and pinning fields, with the average energy computed according to equation (35) in the absence of them, for those instances for which convergence is obtained. Notice that this requires to compute the *soft* fields η_{ka} and ξ_{ka} in addition to the hard ones, which we have done for the sake of the comparison.

In Figure 7 we show the results of this comparison: the addition of reinforcement and pinning fields causes a systematic but very small increase in the average energy per advertiser. In order to quantify the relevance of this effect, its magnitude should be compared to the average *revenue* per advertiser rather than to the value of the energy. We find that the relative effect on the revenue is always smaller than 3×10^{-3} , with an average value and a standard deviation of $3 \times 10^{-4} \pm 4 \times 10^{-4}$.

B. Stability of the replica symmetric solution and phase diagram of the problem

As we have seen in the discussion of paragraph IV A, there is an interval of values of the parameter of the budget distribution \bar{B} for which the algorithm (in the absence of reinforcement) does not converge, and this suggests the

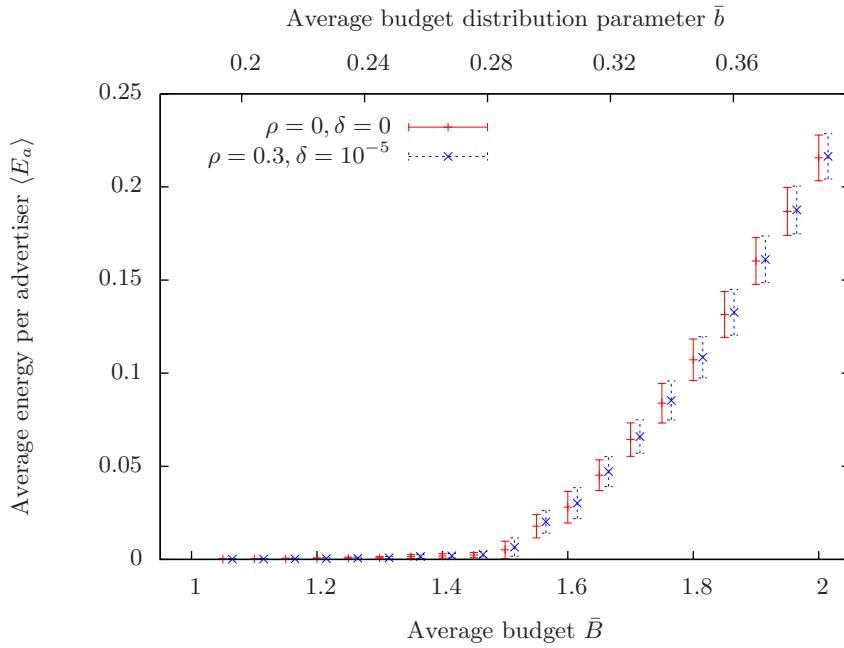


FIG. 7: Comparison of the average energy per advertiser obtained with reinforcement and pinning fields to that obtained without them. The instances are the subset of those analyzed in Figure 5 for which the algorithm without reinforcement and pinning field converges; their total number is 4914. The points corresponding to $\rho = 0.3, \delta_{\max} = 10^{-5}$ have been slightly shifted horizontally to make the comparison possible.

presence of a replica symmetry broken phase. In this paragraph, we are going to present a method which permits to verify this hypothesis directly, and to present some numerical results confirming it. This method was originally introduced in [12] to verify the stability of 1-step RSB (1RSB) solutions, but it is easily simplified to verify the stability of RS solutions as well, as was done for instance in [13, 14].

The idea behind the method is very simple. The RS solution is a special case of a more general solution, the 1RSB solution. The difference between the two is the following: while in the RS solution the value of the cavity fields h_{ka} at the fixed point are well defined, in the 1RSB solution they are defined only *in distribution*. The correct parametrization of the 1RSB solution is not given by a set of *numbers* $\{h_{ka}\}$, but by a set of *functions* $\{\mathcal{P}_{ka}(h_{ka})\}$ which describe the probability distribution of the values of h_{ka} . The RS solution is a special case of this, in which all the functions $\mathcal{P}_{ka}(\cdot)$ are delta functions.

A necessary condition for the RS solution to be correct is that it is *stable* under the enlargement of the solution space to the 1RSB case. This can be verified by considering what happens under the iteration (15, 17) if instead of a sharply defined value the cavity fields h_{ka} have some small but finite variance $v_{ka}/2$ around their fixed point values h_{ka}^* , i.e. if they have a distribution

$$\mathcal{P}_{ka}(h_{ka}) = \mathcal{N}_{ka} e^{-(h_{ka} - h_{ka}^*)^2/v_{ka}} \quad (51)$$

where \mathcal{N}_{ka} is a normalization. Under the iteration (15, 17), the field h_{jb} is computed as a function \mathcal{H}_{jb} of the fields h_{ka} that are at distance 2 from h_{jb} on the graph representing the instance, i.e.

$$\mathcal{H}_{jb}(\{h_{ka}\}) = \mathcal{H}_{jb}(\{\mathcal{U}_{kb}(h_{k \setminus b}) | k \in \partial b \setminus j\}) \quad (52)$$

If $\{h_{ka}\}$ are close to the fixed point values $\{h_{ka}^*\}$, we can expand

$$h_{jb} = \mathcal{H}_{jb}(\{h_{ka}^*\}) + \sum_{k \in \partial b \setminus j} \sum_{a \in \partial k \setminus b} \frac{\partial \mathcal{H}_{jb}(\{h_{ka}^*\})}{\partial h_{ka}} (h_{ka} - h_{ka}^*) + \dots \quad (53)$$

where $\mathcal{H}_{jb}(\{h_{ka}^*\}) = h_{jb}^*$. The probability distribution of the values of h_{jb} is then given by

$$\mathcal{P}_{jb}(h_{jb}) = \mathbb{P} \left[h_{jb}^* + \sum_{k \in \partial b \setminus j} \sum_{a \in \partial k \setminus b} \frac{\partial \mathcal{H}_{jb}(\{h_{ka}^*\})}{\partial h_{ka}} (h_{ka} - h_{ka}^*) = h_{jb} \right] \quad (54)$$

which can be computed by writing the integral representation of the delta function representing this probability and performing twice a gaussian integration. This gives

$$\mathcal{P}_{jb}(h_{jb}) = \mathcal{N}_{jb} e^{-(h_{jb} - h_{jb}^*)^2/v_{jb}} \quad (55)$$

where we identify

$$v_{jb} = \sum_{k \in \partial b \setminus j} \sum_{a \in \partial k \setminus b} \left(\frac{\partial \mathcal{H}_{jb}(\{h_{ka}^*\})}{\partial h_{ka}} \right)^2 v_{ka} . \quad (56)$$

In order to verify the stability of the RS solution towards 1-step replica symmetry breaking, one can then associate a real number v_{ka} to each cavity field h_{ka} , representing the variance of its distribution, and update the values of v_{ka} with equation (56) each time that the field h_{ka} is updated. As the cavity fields reach their fixed point values $\{h_{ka}^*\}$, one can perform one more iteration for all the edges $\{(ka)\}$ in the graph and compare the values of $\{v_{ka}\}$ before the iteration with those $\{v'_{ka}\}$ after it, and measure the *stability parameter* λ :

$$\lambda \equiv \frac{\sum_{k,a} v'_{ka}}{\sum_{k,a} v_{ka}} . \quad (57)$$

If $\lambda > 1$ the RS solution will be unstable towards a 1RSB enlargement of the solution space, otherwise it will be stable.

Notice that, in general, the derivative $\partial \mathcal{H}_{jb} / \partial h_{ka}$ will involve also the *soft* field $\{\xi_{ka}\}$, which therefore must also be computed. This can be seen by writing explicitly

$$\left(\frac{\partial \mathcal{H}_{jb}}{\partial h_{ka}} \right)^2 = \sum_{k \in \partial b \setminus j} \left(\frac{d\mathcal{H}_{jb}}{du_{kb}} \right)^2 \sum_{a \in \partial k \setminus b} \left(\frac{d\mathcal{U}_{kb}}{dh_{ka}} \right)^2 \quad (58)$$

$$= \sum_{k \in \partial b \setminus j} \left\{ \frac{\sum_{x_{b \setminus j} \in x_{b \setminus j}^*(B_b)} x_{kb} e^{x_{b \setminus j} \cdot \eta_{b \setminus j}}}{\sum_{x_{b \setminus j} \in x_{b \setminus j}^*(B_b)} e^{x_{b \setminus j} \cdot \eta_{b \setminus j}}} - \frac{\sum_{x_{b \setminus j} \in x_{b \setminus j}^*(B_b - w_{jb})} x_{kb} e^{x_{b \setminus j} \cdot \eta_{b \setminus j}}}{\sum_{x_{b \setminus j} \in x_{b \setminus j}^*(B_b - w_{jb})} e^{x_{b \setminus j} \cdot \eta_{b \setminus j}}} \right\}^2 \times \\ \times \sum_{a \in \partial k \setminus b} \mathbb{1} \left[h_{ka} = \max_{c \in \partial k \setminus b} h_{kc} \right] \quad (59)$$

where the soft fields $\eta_{b \setminus j}$ are computed as functions of $\xi_{k \setminus b}$ from (16), and where $x_{b \setminus j}^*(z)$ is the set of partial configurations that minimize the function $g(z, x_{b \setminus j})$ defined in (20).

This procedure can be applied to any given instance. Alternatively, it can be used to verify if the RS solution is stable in the thermodynamic limit for the ensemble $\mathcal{E}_\infty(c_a, c_k, \bar{b})$, using the *population dynamics* technique: in this case, the topology of the problem is not fixed, but is extracted at random for each update. One then considers a population (i.e. a set) of N triples (h_i, ξ_i, v_i) . At each step, a local tree of depth 2 is extracted, with average connectivity c_a for the root and c_k for the nodes at depth 1. Each leaf of the tree is associated at random with a triple (h_i, ξ_i, v_i) from the population. The value of the fields h_0 and ξ_0 at the root are computed according to (15–18), and the value of v_0 is computed according to (56). Then, a triple (h_i, ξ_i, v_i) is selected at random from the population and is replaced with (h_0, ξ_0, v_0) . This procedure is repeated a number of times $T_0 \equiv t_0 N \gg N$ (the population size) in order to reach the equilibrium distribution of fields $\mathcal{P}(h)$. The population of variances is then reinitialized to $v_{ka} = 1 \forall (ka)$, and another $T_1 \equiv t_1 N \gg N$ iterations are performed (renormalizing the variances after each $\Delta T = N$ iterations) so that the distribution of variances also equilibrates. Finally, $T_2 \equiv t_2 N \gg N$ iterations are performed, computing λ_T according to (57), and averaging the results. Another quantity which is of interest to characterize the stability of the RS solution is the fraction of non-zero variances v_{ka} in the population, which we shall denote by ϕ . It can be computed together with λ , and it must be zero for the RS solution to be stable.

Using this approach, we have obtained the results shown in Figure 8. The parameters of the distribution of instances are the same as for the data shown in Figures 1: average connectivity $c_a = 10$ for the advertisers and $c_k = 10/3$ for the keywords; the population size is 10^4 ; the equilibration time is $t_0 = 1500$ for the distribution of fields and $t_1 = 1000$ for the distribution of variances; the measurement time is $t_2 = 4000$. We see that the region $\bar{b} \in [0.268, 0.337]$ (corresponding to $\bar{B} \in [1.41, 1.75]$) is unstable, i.e. has $\lambda > 1$. This is in very good agreement with the convergence results of the algorithm, which show a marked drop in the fraction of instances that converge in the same region. Figure 9 shows the corresponding average energy per advertiser and the average entropy per variable. The nature of the two transitions appears different, and it may be interesting to study with more depth how “sharp” is the one at $B = 1.75$.

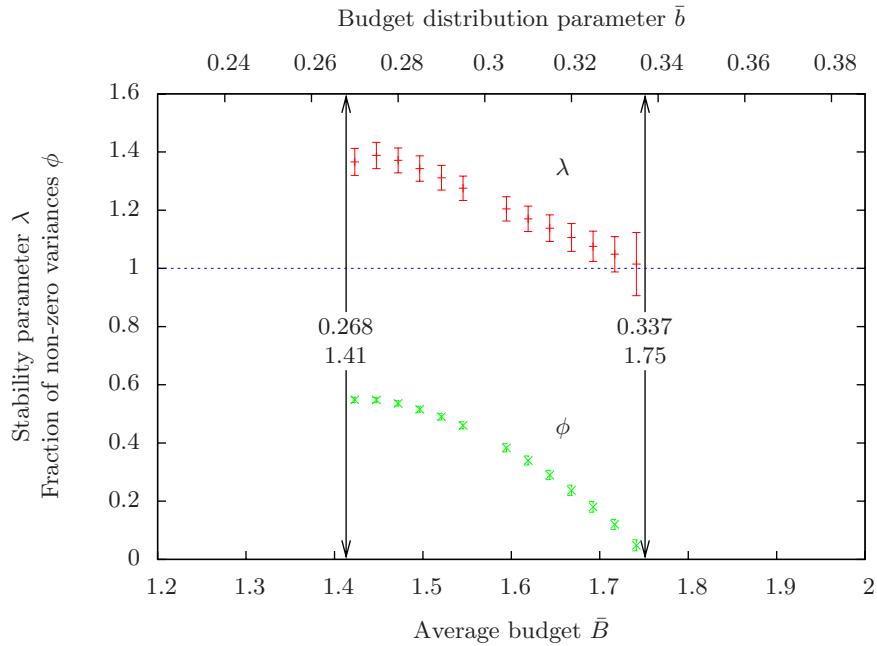


FIG. 8: Stability parameter λ and fraction of non-zero variances ϕ for different values of the budget distribution parameter \bar{b} , obtained from the population dynamics technique. The average connectivities used are the same as in the previous plots, i.e. $c_a = 10$ and $c_k = 10/3$. The region $0.268 \lesssim \bar{b} \lesssim 0.337$ (corresponding to $1.41 \lesssim \bar{B} \lesssim 1.75$) is unstable, i.e. has $\lambda > 1$. Outside of this region, both λ and ϕ vanish. The values of λ in the unstable region are well fitted by a line (except the first point), while the values of ϕ are well fitted by a parabola.

The previous analysis confirms the phase diagram suggested in paragraph IV A: for $\bar{b} < \bar{b}_0$ we have a RS phase with negligible average energy and positive entropy, for $\bar{b}_0 < \bar{b} < \bar{b}_1$ we have a RSB phase, and for $\bar{b}_1 < \bar{b}$ we have again a RS phase, but with positive energy and negligible entropy. The threshold values \bar{b}_0 and \bar{b}_1 depend on the average connectivities c_a and c_k .

C. Comparison to other algorithms

We have compared zero-temperature BP to several other algorithms in terms of the average energy of the solutions they find. The alternative algorithms considered are Simulated Annealing (SA) and Linear Programming (LP) relaxation, with several rounding procedures. In the following paragraphs we shall briefly describe them, and then show and discuss the numerical results we obtained.

1. Simulated annealing

Simulated annealing is an optimization technique first proposed by Kirkpatrick *et al.* [15] which is widely applied in the field of combinatorial optimization and which is inspired by the annealing of solids. It consists in a Metropolis procedure in which the temperature is gradually decreased. In the case of the off-line AdWords problem it is described in Algorithm IV.1. The inputs are an instance of the problem \mathcal{I} , an increasing sequence of inverse temperatures $\{\beta_1, \dots, \beta_n\}$ with $\beta_1 \ll 1$, $\beta_n \gg 1$ and $n \gg 1$, and a large positive integer $T \gg 1$.

The values of T and of $\{\beta_i\}$ must be such that the system thermalizes at any temperature β_i . In the process we can keep track of the configuration with the smallest energy explored in the sampling process, which would be the best approximation for the optimization problem we want to solve. Additionally, keeping track of the average energy for different values of temperature, it is possible to obtain the entropy of the system as $S = \int dE \beta(E)$.

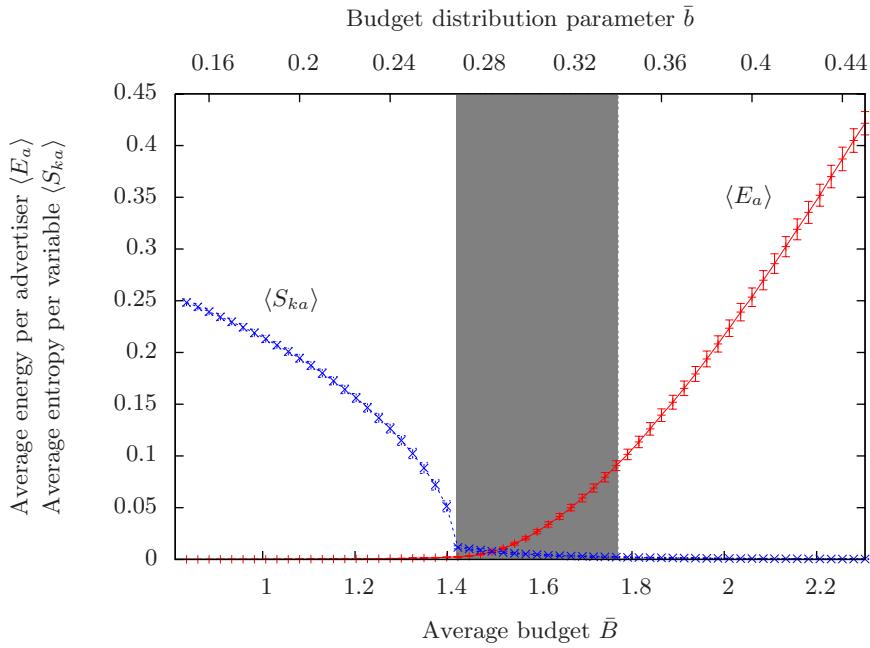


FIG. 9: Average energy per advertiser and average entropy per variable computed with the population dynamics method. The average connectivities used are the same as in the previous plots, i.e. $c_a = 10$ and $c_k = 10/3$. The shaded region $\bar{b} \in [0.268, 0.337]$ corresponds to the RSB phase: in this region, the values obtained for $\langle E_a \rangle$ and $\langle S_{ka} \rangle$ in the RS approximation are (probably) not correct.

Algorithm IV.1: SIMULATED ANNEALING($\mathcal{I}, \{\beta_i\}, T$)

```

for each  $k \in K$ 
  do  $\left\{ \begin{array}{l} \text{select uniformly at random } a \in \partial k \\ \text{for each } b \in \partial k \\ \text{do } x_{kb}^0 \leftarrow \delta_{a,b} \end{array} \right.$ 
   $x_{\min} \leftarrow x^0$ ,  $E_{\min} \leftarrow E(x_{\min})$ 
   $t \leftarrow 0$ ,  $i \leftarrow 1$ 
  while ( $i < n$ )
    while ( $t < iT$ )
       $x' \leftarrow x^t$ 
      select uniformly at random  $k \in K$ 
      select uniformly at random  $b \in \partial k \setminus a$  with  $a : x_{ka}^t = 1$ 
       $x'_{ka} \leftarrow 0$ ,  $x'_{kb} \leftarrow 1$ 
       $\Delta E \leftarrow E(x') - E(x^t)$ 
       $P \leftarrow \min[1, \exp(-\beta_i \Delta E)]$ 
      select uniformly at random  $p \in [0, 1]$ 
      if ( $p < P$ )
        then  $\left\{ \begin{array}{l} x^{t+1} \leftarrow x' \\ \text{if } (E(x^{t+1}) < E_{\min}) \\ \text{then } x_{\min} \leftarrow x^{t+1}, E_{\min} \leftarrow E(x_{\min}) \end{array} \right.$ 
      else  $x^{t+1} \leftarrow x^t$ 
       $t \leftarrow t + 1$ 
     $i \leftarrow i + 1$ 
  return ( $x_{\min}$ )

```

Linear programming (LP) is a common technique for optimization of linear functions over continuous variables. The variables are restricted to a given domain, usually called the *feasible set*, which is specified by a set of constraints that must be satisfied simultaneously by all the variables [16]. Even though LP is defined on continuous variables, it is possible to exploit it in the design of approximation algorithms for problems in combinatorial optimization. The heuristics works as follows: first *relax* all the discrete variables in the original combinatorial problem by allowing them to take continuous values. Then use LP to find an optimal set of values for the continuous variables. In general, the optimal value for the cost function of the *relaxed* problem yields a lower (upper) bound for the original minimization (maximization) combinatorial problem, since the search space is enlarged by the relaxation. Moreover, if it turns out that the solution of the relaxed problem corresponds to *all* the variables taking integer values, then this is a solution for the discrete problem too. Otherwise, a strategy for rounding the continuous variables to integers must be specified. In general, the solution for the relaxation yields a mixture of discrete and fractional variables; it is important to bear in mind that the discrete values obtained in the solution of the LP relaxation do not necessarily correspond to an optimal solution of the original problem. Hence, in general, the rounding scheme should take into account not only the variables with fractional values, but the whole set of them instead. Figure 10 describes an example of integer variables that can be assigned by LP to the relaxation of the AdWords problem which are incompatible with any optimal allocation.

By introducing ancillary *discount* variables $\{d_a\}$ representing the discount d_a given by the auctioneer to advertiser a in case the actual assignment exceeds the budget B_a , the AdWords problem can be restated as follows

$$\max \sum_a \left(\sum_k x_{ka} w_{ka} - d_a \right), \quad (60)$$

subject to

$$\sum_a x_{ka} \leq 1, \quad (61)$$

$$\sum_k x_{ka} w_{ka} - d_a \leq B_a, \quad (62)$$

$$x_{ka} \in \{0, 1\}, \quad (63)$$

which is now a *mixed* integer programming problem, as the discount variables are allowed to take any continuous value. With this formulation, the LP relaxation consists in replacing the integer constraints in (63) with

$$0 \leq x_{ka}^r, \quad (64)$$

where the superscript r stands for *relaxation*. This inequality, along with (61), guarantees that $x_{ka}^r \leq 1$. Moreover, the discount variables d_a are irrelevant in the relaxed version because the budget constraints (62) can be met by tuning the continuous variables x_{ka}^r themselves; in fact they all take value zero in the solution attained by LP. They will play an important role, though, in the rounding process as described below.

A simple way to round the fractional solutions x_{ka}^r given by the LP relaxation is the following *Random Rounding* (RR): assign each keyword k at random to advertiser a with probability proportional to x_{ka}^r . Theorems 4 and 5 of [2] prove that the approximation ratio of RR is $e/(e-1) \simeq 1.582$. This means that if R_0 is the optimal revenue for the problem considered, the solution given by RR will have a revenue R satisfying the inequality

$$R < R_0 < \frac{e}{e-1} R. \quad (65)$$

3. Linear programming with Garg-Kumar-Pandit (GKP) rounding

Garg, Kumar and Pandit [17] proposed another rounding algorithm for the solution of the LP relaxation of the problem. The GKP algorithm takes the integer values obtained by LP as part of the solution, and develops a strategy to round up the remaining variables with fractional values. Theorem 3 in [17] proves that the algorithm to be described has an approximation ratio of $(1 + \sqrt{5})/2 \simeq 1.62$, i.e.

$$R < R_0 < \frac{1 + \sqrt{5}}{2} R, \quad (66)$$

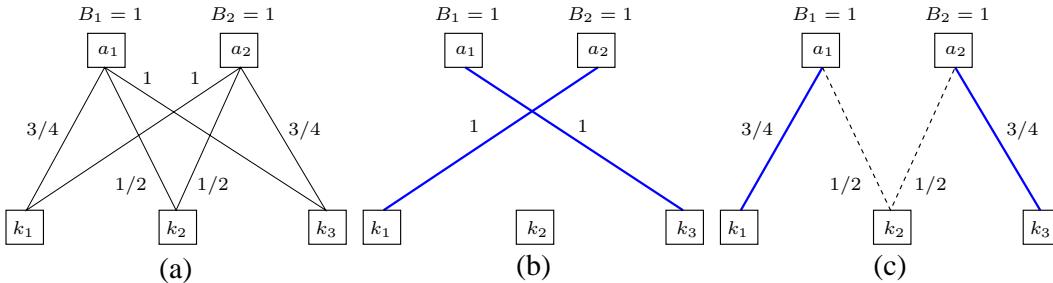


FIG. 10: An example of integer variables that can be assigned by LP to the *relaxed* problem which are not compatible with any optimal solution: (a) Two advertisers bidding for three keywords with budget $B_a = 1$ each; advertiser a_1 bids $3/4$, $1/2$ and 1 for keywords k_1 , k_2 and k_3 , respectively, while advertiser a_2 makes the same bids but in reverse order. (b) An *integer* optimal allocation with the budgets of both advertisers saturated; the missing links correspond to variables assigned to zero, except for keyword k_2 which can be assigned to whichever advertiser with no change in the total revenue. (c) A *fractional* optimal allocation, where the thick blue links correspond to the variables assigned to one, the missing links to those assigned to zero, and the dashed links to the fractional variables, which are equal to $1/2$ in this case. Notice that keyword k_2 is needed in order to saturate the budgets of both advertisers.

with the same notation as in (65). Notice that even though this bound is less favorable than (65), the average case performance of GKP on some distribution of instances could actually be better than RR.

The rounding part of the algorithm is best described by introducing the *residual graph*, which is the graph obtained by removing from the original one all the edges whose associated variables were assigned an integer value by LP. The algorithm applies a series of transformations on the residual graph, consisting of a suitable redistribution of the *mass* on the fractional variables. In this way, without decreasing the degree of optimality reached nor violating any constraint, the residual graph is left with certain special properties; among other things, it does not contain cycles and, most importantly, it is guaranteed that any maximal path⁵ in it starts with an advertiser whose budget is saturated. If the (fractional) amount $x_{ka}^r w_{ka}$ spent by such an advertiser a on the only keyword k linked to him in the residual graph is less than a given fraction p of the budget, i.e. $x_{ka}^r w_{ka} < p B_a$, the corresponding variable is rounded to zero; otherwise the corresponding discount variable d_a is increased, opening up the possibility for the advertiser to acquire such a keyword during the remaining process [17]. Even though the rigorous bound (66) is proved for $p = (3 - \sqrt{5})/2 \simeq 0.38$, for the ensemble studied here better results were observed by using smaller values of p .

4. Linear programming with BP-based rounding

Another alternative to round up the fractional variables is to use the BP algorithm to calculate the minimum energy, under the constraint that the integer variables of the LP relaxation are forced to take the actual integer values given by LP.

5. Numerical results

In order to compare these algorithms to zero-temperature BP, we have generated 300 instances from the distribution $\mathcal{E}(N_a, N_k, N_e, \bar{b})$ with $N_a = 1000$, $N_k = 3000$ and $N_e \simeq 10000$ and with different values of the budget distribution parameter \bar{b} and we have solved each of them with all the algorithms considered. We have grouped the instances in bins according to the value of the average budget \bar{B} , and we have computed the average energy of the solutions found by each algorithm in each bin. The results of the comparison are shown in Figure 11.

The performance of BP and SA relative to the algorithms based on LP rounding appears clearly superior. In particular, notice that the performance of BP-based rounding is much worse than that of BP (especially for low budgets): the solution of the LP-relaxation typically contains some *integer* variables that have the “wrong” value for the integer problem. An example of this can be seen in Figure 10. This suggests that *any* LP-rounding algorithm which assumes that the integer variables in the solution of the relaxation are correct will find a sub-optimal solution

⁵ A *maximal path* is a path that cannot be extended because there are no more edges in the graph contiguous to it.

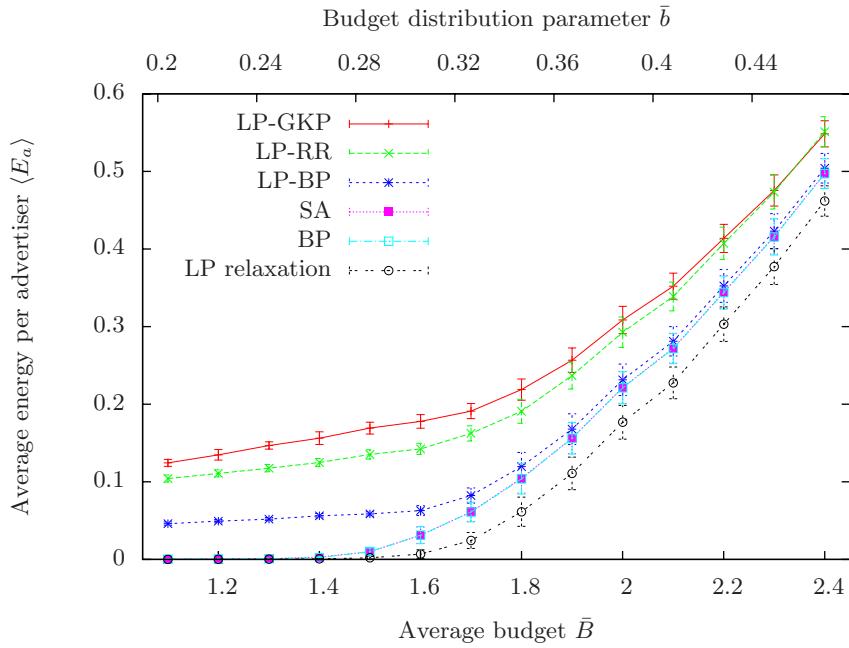


FIG. 11: Comparison of the average energy per advertiser $\langle E_a \rangle$ obtained by the zero-temperature BP algorithm with those obtained by alternative optimization techniques, namely: LP with random rounding (LP-RR), LP with Garg-Kumar-Pandit (LP-GKP) rounding, LP with BP-guided rounding (LP-BP), and simulated annealing (SA). The solution of the relaxed linear program provides a rigorous lower bound for the energy. The curves for SA and BP almost coincide in all the range. The same set of instances with $N_a = 1000$ advertisers, $N_k = 3000$ keywords, and average keyword connectivity $c_k = 10/3$ (which yields on average $\bar{N}_e = 10000$ variables) has been used for all the algorithms.

to the integer problem. This situation is very different compared to what happens for random b-matchings, a similar problem where LP and BP both give exact solutions [18].

The differences in the average energies obtained by BP and SA are always very close to zero, and since the two algorithms are completely different, this suggests that they are also very close to the true optimum. This conjecture is also supported by the analysis of the results for the entropy (not shown): as the equilibration time is increased, the entropy obtained by SA gets closer and closer to the value obtained by BP, suggesting that the latter is in fact exact, which in turn implies that the energy is exact too. Finally, the results of population dynamics for $\bar{B} \notin [1.41, 1.75]$, which gives the average value (over the distribution \mathcal{E}_∞) of the actual optimal energy per advertiser, are also in excellent agreement with the energy obtained with BP and SA, confirming that they are exact at least in this range of values of \bar{B} .

It is worth noting that even though the performances of BP and SA are very similar in terms of the energy of the configurations they obtain, the two algorithms are very different in terms of running time. The running time that SA requires to find a solution with energy $E_{SA} \leq (1 + \epsilon)E_{BP}$ is up to 100 times larger than the running time of BP for $N = 1000$ and $\epsilon = 0.01$.

V. CONCLUSIONS AND PERSPECTIVES

We have described a simple and efficient algorithm which is capable of finding (nearly) optimal assignments for random instances of auctions with budget constraints, based on a statistical mechanics analysis of the problem, and which is equivalent to the zero-temperature Belief Propagation procedure. We have characterized its average performance on a natural distribution of random instances and compared it to other previously studied algorithms, obtaining that it is superior to them in terms of the approximation to the true optimum and/or in terms of running time. We have also studied the properties of the instances distribution in the infinite size limit, deriving a phase diagram (as a function of the parameter \bar{b} of the distribution) with a replica symmetry breaking phase separating two replica symmetric phases, and obtaining a reference for the average value of the true optimum (in the RS phase) against which to test our algorithm.

In the future it would be extremely interesting to test our algorithm on real data (for example of actual bids for internet search keyword auctions), or on a distribution of random instances resembling the actual one more closely.

It would also be interesting to study the generalization of this off-line algorithm to the on-line problem, again from the point of view of the average performance on some distribution of search sequences (possibly time-dependent).

Acknowledgements

We are grateful to C. Mathieu for introducing the problem to one of us, and to M. Mézard, G. Parisi, A. Ramezanpour and V. Pandit for stimulating discussions. This work was partially supported by the Microsoft External Research Initiative.

- [1] B. Lehmann, D. Lehmann and N. Nisan, Combinatorial auctions with decreasing marginal utilities. In *Proceedings of the 3rd ACM Conference on Electronic Commerce*, 18 (2001).
- [2] N. Andelman and Y. Mansour, Auctions with budget constraints. In *9th Scandinavian Workshop on Algorithm Theory (SWAT)*, Springer, 26 (2004).
- [3] Y. Azar, B. Birnbaum, A.R. Karlin, C. Mathieu and C.T Nguyen, Improved approximation algorithms for budgeted allocations. In *Lecture Notes in Computer Science*, Springer **5125** 186 (2008).
- [4] A. Mehta, A. Saberi, U. Vazirani, V. Vazirani, Adwords and generalized online matching, *J. ACM* **54-5** 22 (2007).
- [5] T. Galla, M. Leone, M. Marsili, M. Sellitto, M. Weigt, R. Zecchina, Statistical mechanics of combinatorial auctions, *Phys. Rev. Lett.* **97** 128701 (2006)
- [6] F.R. Kschischang, B.J. Frey and H.-A. Loeliger, Factor graphs and the sum-product algorithm. *IEEE Trans. Infor. Theory* **47** 498 (2002).
- [7] M. Mézard and G. Parisi, The Bethe lattice spin glass revisited. *Eur. Phys. J. B* **20** 217 (2001).
- [8] M. Mézard and G. Parisi, The cavity method at zero temperature. *J. Stat. Phys.* **111** 1/2 1 (2003).
- [9] A. Braunstein, M. Mézard and R. Zecchina, Survey propagation: an algorithm for satisfiability, *Random Structures and Algorithms* **27** 201-226 (2005).
- [10] M. Mézard and R. Zecchina The random K-satisfiability problem: from an analytic solution to an efficient algorithm, *Phys. Rev. E* **66** 056126 (2002).
- [11] A. Braunstein and R. Zecchina, Learning by message passing in networks of discrete synapses, *Phys. Rev. Lett.* **96** 030201 (2006)
- [12] A. Montanari, G. Parisi and F. Ricci-Tersenghi, Instability of one-step replica-symmetry-broken phase in satisfiability problems. *J. Phys. A* **37** 2073 (2004)
- [13] E. Marinari and G. Semerjian, On the number of circuits in random graphs. *J. Stat. Mech. Th. Exp.* **6**:P06019 (2006)
- [14] J. Raymond, A. Sportiello and L. Zdeborová, The phase diagram of 1-in-3 satisfiability problem. *Phys. Rev. E* **76** 011101 (2007)
- [15] S. Kirkpatrick, C.D. Gelatt Jr, and M.P. Vecchi, Optimization by simulated annealing, IBM Research Report RC 9355 (1982). S. Kirkpatrick, C.D. Gelatt Jr, and M.P. Vecchi, Optimization by simulated annealing, *Science* **220** 671 (1983).
- [16] J., Matoušek and B. Gärtner, *Understanding and using linear programming*, Springer (2007).
- [17] R. Garg, V. Kumar and V. Pandit, Approximation algorithms for budget-constrained auctions. In *Lecture Notes on Computer Science*, Springer **2129** 102 (2001).
- [18] M. Bayati, C. Borgs, J. Chayes, R. Zecchina, On the exactness of the cavity method for weighted b-matchings on arbitrary graphs and its relation to Linear Programs, *J. Stat. Mech.* L06001 (2008)